Working Paper in Economics No. 706

Reducing the gap between stated and real behavior in transportation studies: The use of an oath script

Fredrik Carlsson, Elina Lampi and Hang Yin

Department of Economics, September 2017

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Reducing the gap between stated and real behavior in transportation studies: The use of an oath script

Fredrik Carlsson,[a] Elina Lampi,[b] Hang Yin[c]

Abstract

We investigate whether taking an oath in a survey situation affects respondent behavior in choice experiments with a focus on travel time, comfort, and cost. We conduct two studies in Beijing: one with car commuters and one with public transport commuters. Overall, we find little difference in behavior between responses with and without an oath. In fact, the only difference is that responses are more internally consistent in the version with an oath script: the respondents trade off money in terms of a fuel cost and in terms of a congestion charge in the same way. However, there is no statistically significant effect on marginal willingness to pay for any of the attributes.

Key-words: hypothetical bias, oath, stated preferences

JEL- classification: C80, D03, R41

[a] Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden, fredrik.carlsson@economics.gu.se
[b] Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden, elina.lampi@economics.gu.se
[c] (corresponding author) Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden, hang.yin@economics.gu.se

1. Introduction

The use of stated preference (SP) methods has a long tradition in transport economics. There are many reasons for why SP methods are used, but perhaps the most important is that they are the only methods that in a convincing way can measure a value of a non-existing good or of a change in the quality of an existing good. They are also the only methods that can measure non-use values in a direct way. At the same time, the main criticism against SP methods is that they measure *stated* choices/behavior and not *actual* choices/behavior observed in the market. The difference between stated willingness to pay (WTP) and the corresponding WTP if the payment were for real is often denoted hypothetical bias in the literature. Public goods and non-use values are the major explanations of why hypothetical bias occurs (List and Gallet, 2001). Although the findings are mixed, there is considerable evidence that WTP estimates are higher in SP than in the corresponding real setting; see for example Cummings et al. (1995), Cummings et al. (1997), and Frykblom (1997), and as well as List and Gallet (2001) and Murphy et al. (2005) for meta analyses. In the area of transport, there are also a few studies investigating the difference between stated and real choices/WTP. Hensher (2010) investigated the difference between revealed and stated preferences captured in three different choice experiment (CE) studies on transport mode and ticket type choices and found no evidence of hypothetical bias. Swärdh (2008) found a weak but not statistically significant tendency of positive hypothetical bias in WTP for travel time reductions. However, using information from a follow-up certainty question, he found that this hypothetical bias was only evident for non-certain subjects. Fifer et al. (2014) found a positive hypothetical bias comparing stated and revealed preferences for a kilometer-based charging scheme. One interesting finding in transport is that a set of studies actually find a negative hypothetical bias for value of time, i.e. that people state a lower WTP in a study than what they would pay for time savings in a real situation; see Brownstone and Small (2005), Isacsson (2007), and Hultkrantz and Xue (2010). There are several explanations for this perhaps surprising finding. One is that people have the wrong perception of travel time, in particular when they face unexpected delays, e.g., a delay of five minutes could be perceived to be much longer. However, in a stated preference setting, five minutes would be perceived as five minutes. Another explanation could be that time constraints are not fully attended to in a stated preference setting, i.e., respondents do not carefully enough consider what their alternative use of time is.

If we disregard studies looking at the environmental impacts of transport, it is fair to say that SP studies in transport economics to a much lesser extent involve public goods and non-use values. Furthermore, the main interest in many of the studies in transport economics is in estimating marginal trade-offs among the attributes, often the marginal value of time. Estimates of marginal trade-offs are considered to be less prone to differences between stated and real behavior (Carson and Groves, 2007). We should perhaps therefore expect the discrepancy between stated and real behavior to be lower in SP studies in transport economics than in fields such as environmental and health economics. However, one type of SP study in transport is likely to trigger rather strong attitudes and thus implies a higher risk of ending up with a difference between stated and real behavior: those that involve infrastructure investment, and in particular congestion charge systems; see Li and Hensher (2012) for an overview of stated preference studies on congestion charges and car use.

Several strategies have been developed to reduce the hypothetical bias in SP studies, including: (i) follow-up certainty questions (Champ et al., 1997; Champ and Bishop, 2001), (ii) cheap talk scripts (Cummings and Taylor, 1999; Carlsson et al., 2005), (iii) consequential scripts (Bulte et al., 2005; Vossler and Evans, 2009; Barrage and Lee, 2010), and (iv) time-to-think protocols (Cook et al., 2007; Whittington et al., 1992). The cheap talk script is probably the most frequently used one. The underlying idea of the script is that, by informing respondents about hypothetical bias, they become less prone to it. The observed effect of the cheap talk script varies among studies and its success seems to depend on the characteristics of the good, the length of the script, and the valuation method (see Aadland and Caplan, 2003, 2006; Carlsson et al., 2005; Cummings and Taylor, 1999; List, 2001; and Murphy et al., 2005). A more recent approach for addressing hypothetical bias is the use of an oath script. Jacquemet et al. (2013), Carlsson et al. (2013), and Stevens et al. (2013), all found lower WTP values in treatments with oath compared with treatments without oath. De-Magistris and Pascucci (2014) is the only study we are aware of that investigate the role of an oath in a choice experiment setting. They found that taking an oath results in lower marginal WTP estimates.[1] With an oath, respondents are asked to swear (or promise) to answer truthfully, mimicking the act of taking an oath in a courtroom. According to social psychology research, it is not enough to only inform people about bad behavior (as is the case with the cheap talk

[1] There are also experimental studies looking at other types of behavior and oath. For example, a recent study by Jacquemet et al. (2015) found that taking an oath increased a person's willingness to tell the truth in an experimental study.

script). In addition to this information, people also need to commit to the task in order to change their behavior (see, e.g., Joule et al., 2007). Thus, promising to answer truthfully or signing an oath before the respondent participates in a stated preference study is an attempt to use commitment theory to reduce hypothetical bias.

We conducted two studies related to commuting: one directed to car commuters and one to public transport commuters. Both studies were conducted in Beijing, China, which is a city with substantial traffic and air quality problems (Creutzig and He, 2009; Hao and Wang, 2005; Westerdahl et al., 2009). The objective of the present paper is to investigate the effect of an oath script on respondent behavior in a transport-related choice experiment with a focus on travel time, travel comfort, and travel cost. Carlsson et al. (2013) found that share of zero WTPs as well as extremely high WTPs decreased among Chinese respondents who took an oath in a contingent valuation (CVM) survey. We are therefore interested in investigating whether individuals who take an oath will make more careful choices and take all available attributes into account.

We are not aware of any other study that has tested whether a commitment in the form of an oath matters for stated choice and WTP in transport-related choice experiments. As for the effect of taking an oath on people's behavior, however, a few studies can be discerned and most of them are experimental studies.. Jacquemet et al. (2013) investigate the impact of an oath script in two laboratory experiments: one induced value second-price auction and one homegrown value second-price auction. Both experiments had three treatments: a hypothetical, a real setting with monetary incentives, and a hypothetical with the option of signing an oath before participating. By signing the oath, subjects agreed to "swear on their honor" to tell the truth and provide honest answers. The authors found that subjects who took the oath were on average less likely to overstate or understate their bids; that is, the variance of bids was reduced. They also found that the hypothetical payment treatment with the oath outperformed both the hypothetical non-oath treatment and the real payment treatment. The authors also discussed possible reasons for why signing an oath helps reduce/correct the hypothetical bias and concluded that an oath increases the willingness to tell the truth due to a strengthening of the intrinsic motivation to do so. A second experimental study by Jacquemet et al. (2017) compared voting behavior in real and hypothetical referenda and found that the oath script eliminated hypothetical bias. A third experimental study by the same authors (Jacquemet et al., 2014) investigated the effect of an oath script in a classic coordination game

(with senders and receivers) in a laboratory setting. Here the authors found that taking an oath increases optimal coordination by nearly 50 percent. Senders were more likely to send informative and reassuring messages and then do what they said they would. More receivers trusted the messages received under an oath than without it. Finally, the fourth experimental study is Stevens et al. (2013), who compared WTP across three treatments: two hypothetical that were followed by an actual payment and one with an actual payment only. The difference between the first two treatments was that before the second treatment started, the participants were asked to sign an oath. Their results show that with an oath, mean hypothetical payments did not differ from mean actual payments.

We are aware of only two stated preference studies that investigate the effect of an oath script. The first is Carlsson et al. (2013), who conducted a CVM study about climate change and tested whether taking an oath impacts people's WTP. The study was conducted in both Sweden and China. They found that in both countries, the shares of zero and extremely high WTP responses decreased when an oath script was used, resulting in lower WTP variance. The second is De-Magistris and Pascucci (2014), who found lower marginal WTP in a CE with an oath script compared with using a cheap talk script or no script at all.

The rest of the paper is organized as follows. In Section 2, we describe the two SP studies and the oath script used. Section 3 gives a description of the administration of the studies and of the sample used. We report the results in Section 4 and discuss the implications of our findings in Section 5.

2. Description of the SP studies and the oath script

We investigate the role of an oath script in two different choice experiments. The purpose of the two experiments was to investigate the value of time among car and public transport commuters in Beijing and whether taking an oath affects survey responses. For a more detailed description of the experiments and their results, see Yin (2017).

Both studies started with the subjects responding to questions about individual and household characteristics, as well as a set of questions about commuting by car and public transport. Then we elicited their attitudes to existing transport policies by asking them to rate each

policy on a scale from 1 to 10.[2] After this, the choice experiment was introduced. We started the experiment with a cheap talk script and the oath question (not in the control groups). In a next step, we described the scenario, which differed between the car and public transport studies, and conducted the actual choice experiment. Finally, the respondents were asked a set of follow-up questions about attribute attendance.

2.1 Study 1: A choice experiment for car commuters

A generic choice experiment with 12 choice sets was presented to commuters from households that owned a car. In each choice situation, there were two alternative commuting schemes, both by car. Each alternative was described with four attributes: travel time in the morning, travel time in the afternoon, a congestion charge, and daily fuel cost.

In the present paper, we want to investigate the trade-offs between travel time and cost of driving. It is clear from the pilot studies that, for many commuters, travel time reductions in the morning are much more important than reductions in the afternoon. Therefore, commuting time in the morning and afternoon are included as two separate attributes. Commuting time refers only to the time spent in the vehicle, while the walking time, either from home to the parked car or from the parked car to work, is not included. We utilized a pivot design where the levels of the travel time attributes were based on the current travel time for the respondents. The third attribute, congestion charge, is the daily price paid for driving into the central city during rush hours. This attribute has four levels ranging from 10 to 25 yuan per day. The daily fuel cost is used as the cost attribute as it represents the main expense of driving, alongside parking fees or road toll, and applies to all drivers. The levels of the fuel cost attribute were also pivoted upon real cost. Table 1 presents the attributes and their levels.

2.2 Study 2: A choice experiment for public transport commuters

We designed a generic choice experiment for public transport commuters to elicit their preferences regarding travel time, level of crowding, and travel cost. There were in total 12 choice sets, each with two alternatives describing two different public transport commuting schemes. Again each alternative was described with four attributes: travel time in the morning, travel time in the afternoon, level of crowding, and ticket price.

[2] The existing transportation policies include driving day restriction, higher fuel tax, higher parking fee, a lottery for new vehicle plates, and low-priced public transport. The value 1 on the scale indicates "this policy does not help solve the congestion problem at all" and 10 "this policy is very helpful in solving the congestion problem."

Public transport commuting time referred to the time spent on the bus or subway, or both, and as well as the time spent waiting for a transfer. The time an individual spent on the bus and the subway, respectively, was not distinguished, as many respondents use both. The levels of the time attributes were also pivoted upon current individual travel time. The third attribute was level of crowding onboard, expressed as number of days out of ten commuting days experiencing severe crowdedness inside the bus or subway car. We assumed there are only two states, severely crowded and crowded, since both buses and subway are crowded most of the time during peak hours in Beijing. In the survey, all respondents were shown the same two photos of different numbers of passengers standing on board. Thus, we believe that respondents had a consistent view of what we meant by crowded and severely crowded. These photos are shown in the Appendix. The final attribute was the ticket cost of using public transport including buses and subway. The level varies from 4 to 10 yuan per day, which is equal to or higher than the real cost at the time of the study.

2.3 Experimental design for both studies

For both two experiments, we used Ngene to generate a D-optimal design (Hensher, Rose, and Greene, 2005). Prior information about the parameters is required in D-optimal design and was obtained from focus group discussions and four pilot studies of in total 80 observations.[3] The main survey was conducted from May to July 2014.

A D-optimal design is considered more efficient than an orthogonal design since it allows parameters to be estimated with minimum variance and enables us to avoid dominance between alternatives and extract more information from each respondent with fewer choice situations. For both experiments, we generated 12 choice situations with two alternatives in each situation. In both situations, the real travel time varies considerably among the participants. Thus, we have them separated into two groups: a “long” group for those with a travel time of at least 45 minutes in the morning and a “short” group for those with a travel time of less than 45 minutes in the morning. Different time and cost levels are used in the long and short groups.

[3] The focus group study and the four pilot studies were conducted from July to November 2013.

Table 1. Attributes and levels in the two experiments

Attributes	Levels		Interpretation
	Long	Short	
Study 1: Choice experiment for car commuters			
Morning	6,9,12,15	4,6,8,10	x minutes less than real commuting time
Afternoon	6,9,12,15	4,6,8,10	x minutes less than real commuting time
Congestion Charge	10,15,20,25	10,15,20,25	fixed number
Fuel cost	-6,0,8,18	-5,0,5,10	yuan less or more than current level
Study 2: Choice experiment for public transport commuters			
Morning	6,9,12,15	4,6,8,10	x minutes less than real commuting time
Afternoon	6,9,12,15	4,6,8,10	x minutes less than real commuting time
Level of crowding	0	0	6 out of 10 days are severely crowded; otherwise crowded
	1	1	8 out of 10 days are severely crowded; otherwise crowded
	2	2	always severely crowded (today)
Ticket price	4,6,8,10	4,6,8,10	yuan/day

2.5 The oath script

A cheap talk script was read to all respondents before moving on to the section with the choice situations. [4] The respondents were also encouraged to estimate their behavior without any consideration or influence of their opinion about the congestion charge policy.[5] Half of the respondents were randomly selected using a computer-assistant personal interview (CAPI) system (see also Section 3) and then asked to respond to an oath question immediately before the choice experiment. The oath script was formulated as follows: "Do you feel you can promise us you will answer the questions that will follow truthfully?" The available options were "Yes, I promise to answer the questions in the survey truthfully" and "No, I cannot promise." Almost all respondents who received this question answered "yes." In the analysis, we will therefore include all respondents, irrespective of their response to this question.[6]

[4] The cheap talk script read as follows: "Before making your choices, we would like you to consider how the changes will affect you, for example in terms of increased costs. Previous studies of this kind have shown that some people tend to overstate that they are willing to change their behavior very easily, while others tend to overstate that they are very unwilling to change their behavior."

[5] The text aimed to encourage respondents to state their behavior regardless of their opinion about the policy read as follows: "Furthermore, in this part of the survey, we want you to consider how you would change your behavior if the policies were implemented, and not your opinion on what policies should be implemented. So irrespective of whether you, for example, think the congestion charge is a good or bad policy, we want you to make your choices under the assumption that there is a congestion charge."

[6] In the first study, 283 of the 290 respondents who received the oath version of the survey said "yes." In study 2, the corresponding share was 354 out of 358. The main results that we will present are not sensitive to whether or not we include those who said no.

3. Administration of surveys and description of the samples

The transportation network in Beijing is based on six ring roads with the Forbidden City in the center. The area within the third ring road is extremely congested during rush hours, and it is also the area that will most likely be subject to the congestion charge scheme, if implemented. Therefore, our studies focus on commuters who have to travel across the third ring road during rush hour, since they are most likely to be affected by the policy.

The face-to-face interviews were conducted using laptops with a computer-assisted personal interviewing (CAPI) system installed. Thus, we could generate individualized choice situations for each respondent. Twenty-four experienced enumerators were hired to carry out the survey, and this group was provided a training session on the questionnaire and the CAPI system.

To find eligible respondents, we used a stratified sampling method with a screening question. Thirty-five residential areas and five business districts were chosen. Our sample covers commuters living both outside and inside the third ring road.[7] Households were randomly selected from these places. The number of interviewed households is about 30 for each residential area and about 200 for each business district. At the very beginning of each survey, a screening question was used to check the eligibility of the selected households. It read: "Does any adult household member who is at least 18 years old need to enter inside the third ring road during morning/evening rush hours?" If the answer was no, the survey was terminated. If there was more than one eligible commuter in a household, we randomly picked one of them using the CAPI system as described below.

In total, 1,475 commuters were interviewed in the two studies. The number of completed interviews was 1,370 (622 for car owners and 748 for public transport commuters). Twenty-three observations are dropped due to negative time and cost attribute levels.[8] As a result, the samples used in the analysis include 605 car commuters and 742 public transport commuters. In both studies, 48 percent of the respondents received the oath script; see Table 2.

[7] Eleven out of the 35 residential developments were located inside the third ring road. Two of five business districts were located inside the third ring road, while another two business districts were located outside the third ring road.

[8] In our experimental design, the levels of commuting time and fuel cost are pivoted on the current levels. In some cases this resulted in negative time or cost levels in the experiment, since for example a respondent could have a travel time of less than 10 minutes. These respondents are completely dropped from the analysis.

Table 2. Total number of observations

	Oath	No Oath	Total
Study 1: Generic experiment, car commuters	290	315	605
Study 2: Generic experiment, public transport commuters	358	384	742
Total	648	699	1347

Table 3. Descriptive statistics of key variables in the sample

		Oath		No Oath	
		Mean	SD	Mean	SD
Study 1: Choice experiment for car commuters					
Commuting time in the morning	minutes	42.41	16.71	44.12	18.33
Commuting time in the afternoon	minutes	48.14	20.65	49.13	19.31
Commuting time with severe congestion	minutes	60.58	27.37	66.50	29.35
Fuel cost	yuan per day	29.16	26.37	29.11	15.65
Share of respondents with travel time >45 min	>45 minutes	0.38		0.43	
Individual income in year 2013	10,000 yuan	7.60	5.71	7.72	7.37
Share of respondents with high yearly income	>70,000 yuan	0.43		0.45	
Attitude towards congestion charge	0 to 10 scale	5.93	2.19	5.75	2.10
Gender	1=male	0.51	0.50	0.57	0.50
Age	years	36.43	10.41	35.91	10.23
Share of university level education	1=university	0.39	0.49	0.38	0.48
Study 2: Choice experiment for public transport commuters					
Commuting time in the morning	minutes	48.72	22.28	48.28	19.80
Commuting time in the afternoon	minutes	51.47	21.45	51.31	20.24
Commuting time with severe congestion	minutes	58.36	25.05	59.36	25.87
Ticket cost	yuan per day	3.02	1.80	2.98	1.86
Share of respondents with travel time >45 min.	>45 minutes	0.51		0.50	
Individual income in year 2013	10,000 yuan	5.63	3.03	5.66	3.18
Share of respondents with high yearly income	>70,000 yuan	0.25		0.25	
Gender	1=male	0.40	0.49	0.48	0.50
Age	years	34.63	11.35	34.70	10.86
Share of university level education	1=university	0.30	0.46	0.27	0.44

The average commuting time is roughly 40 to 50 minutes for both car drivers and public transport users, and it is a bit longer in the afternoon than in the morning. In case of severe congestion, the travel time can be more than one hour. About 40 percent of the car commuters have to spend 45 minutes or more each direction. The corresponding share for public transport users is about 50 percent. The daily fuel cost for car commuters is about 30 yuan. A daily public transport ticket is considerably cheaper.

As expected, there are some differences between car and public transport commuters. Car commuters spend less time but more money on commuting than do public transport users.

The average income of car users is higher, and this group also has a large share of high-income people. A t-test allows us to compare the two groups with and without oath. For car commuters, there are no statistically significant differences (at the 95% level) between the oath and no oath survey versions for the variables listed in Table 3. For public transport commuters, there are slightly more females in the groups that received the oath (Pearson Chi^2 (1) =5.09 and p-value=0.024).

4. Results

For each of the two experiments, we begin by estimating separate models for the two versions with and without the oath script. In all models, we pool the survey versions with short and long travel time, since these two versions were primarily implemented to improve the design of the experiment. We estimate random parameter logit (RPL) models and report coefficient estimates and marginal willingness to pay (MWTP) estimates. In both the car and the public transport experiment, we keep the cost attribute fixed, and the remaining attribute coefficients are assumed to be normally distributed. We report mean MWTP and the standard deviation of MWTP in order to consider the preference heterogeneity in terms of WTP as well.

MWTPs for the time attributes are measured in yuan per minute. We also estimate an MWTP for the charge attribute, although this attribute is also a monetary measure. We do this primarily to be able to make a direct comparison between the oath and no-oath versions. Since both the congestion charge and fuel cost are monetary attributes and measured in Yuan, we can also conduct a consistency test by investigating whether the ratio between the charge attribute and the cost attribute is one. Moreover, since one of our objectives is to investigate whether those answering an oath version of the survey will make more careful choices than those not answering an oath survey, we also look at the impact of oath on possible attribute non-attendance (ANA). The probabilities of not attending to an attribute in both experiments are estimated using a latent class model (LCM); see, e.g., Scarpa et al. (2009). More specifically, we estimated an LCM model with five classes, where in four of the classes one attribute is not attended to while the others are. All the attributes are attended to in the fifth class. Table A1 in the Appendix reports the probabilities of attending to the attributes according to the LCM. We find that the probabilities of attending to the attributes are very similar in the two versions with and without an oath. Thus, there is no evidence supporting the hypothesis that taking an oath makes respondents more attentive to the attributes. However, we will still focus our analysis on the RPL where we take attribute non-attendance into

account. Thus, using the estimated non-attendance probabilities from the latent class model, we estimate RPL models using 1,000 Halton draws. These are presented in Table 4. We report coefficient estimates but also MWTP estimates, including standard deviations of the MWTP. We also report the p-values from the t-tests $MWTP_{Oath} - MWTP_{no\ oath}$ in Table 4.

Table 4. Random parameter models for the choice experiments. Attributes are weighted based on estimated ANA from an LCM. Standard errors in parentheses.

	Car					Public transport				
Random parameters	No Oath		Oath		$MWTP_{Oath} - MWTP_{No\ Oath}$	No Oath		Oath		$MWTP_{Oath} - MWTP_{No\ Oath}$
	Coefficient	MWTP	Coefficient	MWTP	p-value	Coefficient	MWTP	Coefficient	MWTP	p-value
Time morning	-0.132***	-0.590***	-0.128***	-0.595***	0.951	-0.086***	-0.122***	-0.116***	-0.161***	0.432
	(0.012)	(0.045)	(0.013)	(0.051)		(0.024)	(0.033)	(0.027)	(0.036)	
Time afternoon	-0.048***	-0.214***	-0.057***	-0.265***	0.410	-0.003	-0.005	-0.001	-0.001	0.864
	(0.010)	(0.039)	(0.012)	(0.049)		(0.011)	(0.015)	(0.012)	(0.059)	
Charge	-0.248***	-1.106***	-0.216***	-1.003***	0.183					
	(0.016)	(0.053)	(0.016)	(0.056)						
Comfort						-0.460***	-0.654***	-0.492***	-0.680***	0.863
						(0.077)	(0.104)	(0.084)	(0.109)	
Fixed parameters										
Cost	-0.234***		-0.216***			-0.704***		-0.724***		
	(0.010)		(0.011)			(0.025)		(0.026)		
Standard deviations										
Time morning	0.102***	0.456***	0.111***	0.512***	0.544	0.249***	0.354***	0.290***	0.401***	0.458
	(0.014)	(0.063)	(0.014)	(0.068)		(0.031)	(0.040)	(0.030)	(0.050)	
Time afternoon	0.042**	0.188**	0.093***	0.433***	0.015	0.077***	0.110***	0.084***	0.116***	0.937
	(0.018)	(0.080)	(0.014)	(0.062)		(0.019)	(0.026)	(0.018)	(0.024)	
Charge	0.178***	0.794***	0.175***	0.810***	0.863					0.495
	(0.013)	(0.062)	(0.013)	(0.067)						
Comfort						0.967***	1.374***	1.064***	1.469***	
						(0.074)	(0.096)	(0.081)	(0.100)	
No. of groups	315		290			384		358		
No. of observations	3780		3480			4608		4296		
R2	0.198		0.171			0.411		0.411		

*** = significant at 1% level, **= significant at 5% level, and *= significant at 10% level.

First of all, all the parameters have the expected signs. For both groups, but especially for public transport commuters, travel time reductions are much more important in the morning than in the afternoon. In addition, public transport commuters clearly care about the level of crowding on board buses and subway trains. The coefficients of the standard deviations are significant, revealing that there is unobservable heterogeneity in preferences for the attributes. The largest variations seem to be for the charge attribute in the car sample and for the comfort attribute in the public transport sample.

For car commuters, the MWTPs for travel time reductions are very similar between the two versions, although marginally higher in the version with an oath script. In contrast, the MWTP for the congestion charge is slightly lower in the survey version with an oath script. Using t-tests, we cannot reject the hypothesis of equal MWTPs between the two versions – with and without an oath script – at the 10% level. The estimated standard deviations of MWTP are either almost the same or slightly higher in the oath script sample. The difference is statistically significant in one case: the standard deviation of the MWTP for travel time reduction in the afternoon is significantly larger in the oath version (p-value = 0.015).

In the public transport experiment, the MWTPs are consistently slightly higher in the version with an oath script. However, using t-tests we cannot reject the hypothesis of equal MWTPs between the two versions – with and without an oath script – for any of the attributes in the public transport sample at the 10% level. The estimated standard deviations of MWTP are either almost the same or slightly higher in the oath script sample, but again there are no statistically significant differences.

Consequently, we find basically no evidence supporting the hypothesis that an oath script affects estimated preferences for an attribute; both mean and standard deviation of the estimated MWTP are very similar across the two samples. As discussed above, we can also compare the preferences for charge and fuel cost. Respondents “should” value a one yuan increase in congestion cost the same as a one yuan increase in total fuel cost. Thus, the ratio between the attribute coefficients for charge and cost should be one. The estimated ratios are close to one in both samples. However, there is a small but statistically significant difference between the oath and no oath versions: The MWTP of the charge attribute in the no-oath script sample is statistically significantly higher than one (1.106; t-test p-value = 0.047). Thus, on average, respondents express a stronger preference against an increase in the congestion

charge costs compared with an increase in the fuel price. However, in the version with an oath script, the ratio is only slightly above one and not statistically different from one (1.003; t-test p-value = 0.958).[9] Thus, respondents who answered a survey with an oath script pass our "consistency test" about costs, while those in the version without an oath script do not. Since the congestion charge is the attribute that is potentially sensitive, and linked to transport policy in general, this might not be surprising. It also suggests that the oath script does matter, but only at the margin. However, as can be seen in Table 4, the MWTPs of the charge attribute in the oath and no oath versions are not significantly different from each other (t-test; p-value is 0.183).

5. Discussion

We designed two choice experiments to investigate WTP for a transport-related good in Beijing, China: one for car commuters and one for public transport commuters. In particular, we investigated whether taking an oath affects stated MWTP levels and possible attribute non-attendance in a choice experiment study. Asking respondents to take an oath before participating in a stated preference study is an attempt to use commitment theory to reduce a potential problem with hypothetical bias. We found very little evidence that an oath script affects respondent behavior in the context of a transport-related good. We could not reject the hypothesis of equal MWTPs in the oath and no oath versions at the 10 % significance level. Nor did we find any clear differences in the probability of not attending to the attributes. The only indication of an effect of taking an oath is that it made the respondents consider the costs of fuel and congestion charge in a more consistent way. However, the difference compared with the version without an oath was small. Thus, our study does not confirm previous tests of an oath script (Jacquemet et al., 2010, 2013; Carlsson et al., 2013; Stevens et al., 2013; De-Magistris and Pascucci, 2014, and Qin and Shogren, 2017). Previous research has even found an effect of an oath script on Chinese respondents (Carlsson et al., 2013 and Qin and Shogren, 2017), suggesting there are no cultural explanations for the absence of an effect. It is possible that our way to ask respondents to answer truthfully is too vague. Previous research show however mixed results of the wording of an oath: Qin and Shogren (2017) studied oath in two laboratory experiments in China. They found that a promise to swear upon one´s honor (a promise script) reduced hypothetical bias in one of the experiments, while not in the other one. They then replaced the promise script in the second experiment by a stronger

[9] As shown in Table 4 also the MWTPs of charge in the oath and no oath versions are significantly different from each other; p-value= 0.183.

commitment (a solemn oath) where respondents were asked to *swear* upon one´s honor and found a decrease in hypothetical bias. However, Carlsson et al., (2013) used a promise script in China and Sweden and found an effect on respondent behavior.

The lack of an oath effect in the present study could possibly be attributed to an absence of hypothetical bias. We are unfortunately not able to test this hypothesis, but it is in line with previous findings by Hensher (2010), who in three different CE studies on transport mode and ticket type choices did not find any evidence of hypothetical bias. Similarly, Swärdh (2008) found no hypothetical bias in willingness to pay for travel time reductions among certain subjects. Another hypothesis is that oath scripts simply do not always work, at least not for all kinds of goods. It is difficult to compare our results with previous studies since most of the previous studies have used different methods, such as laboratory experiments (Jacquemet et al., 2010 and 2013; Stevens et al., 2013; and Qin and Shogren, 2017), or CVM: (Carlsson et al., 2013). However, De-Magistris and Pascucci (2014) found that an oath script decreased hypothetical bias in a CE study but their good was very different from ours, namely food (sushi). Thus, despite many positive findings regarding the effects of oaths on hypothetical bias, we cannot claim it to be a universal tool to decrease or mitigate hypothetical bias in all types of SP studies. It is possible that transport-related goods significantly differ from for example environmental goods. More research about the use of oaths in studies on transport-related goods is clearly needed to address their possible impact.

Acknowledgments

Financial support from the Swedish International Development Cooperation Agency (Sida) to the Environmental Economics Unit at the University of Gothenburg, the Sustainable Transport Initiative at the University of Gothenburg, and Chalmers University of Technology, and from the Formas Commons Program, is gratefully acknowledged. We have received valuable comments from Olof Johansson-Stenman, Mitesh Kataria, Eric Verhoef, and Jintao Xu.

References

Aadland, D., and A.J. Caplan, (2003), Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias. *American Journal of Agricultural Economics* 85: 492–502.

Aadland, D., and A.J. Caplan, (2006), Cheap Talk Revisited: New Evidence from CVM. *Journal of Economic Behavior and Organization* 60: 562–578.

Barrage L., and M. Sok Lee (2010). A penny for your thoughts: Inducing truth-telling in stated preference elicitation", *Economic Letters* 106: 140–142.

Brownstone, D., and K.A. Small (2005), Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A* 39: 279–293.

Bulte, E., S. Gerking., J.A. List, A. de Zeeuw. (2005). The effect of varying the causes of environmental problems on stated WTP values: Evidence from a field study. *Journal of Environmental Economics and Management* 49: 330–342.

Carlsson, F. P. Frykblom,, and C. J. Lagerkvist, (2005), Using Cheap-Talk as A Test of Validity in Choice Experiments. *Economics Letters* 89: 147–152.

Carlsson F., M. Kataria, A. Krupnick, E. Lampi, Å. Löfgren, P. Qin, and T. Sterner (2013), The Truth, the Whole Truth, and Nothing but the Truth- A Multiple Country Test of an Oath Script. *Journal of Economic Behavior & Organization* 89: pp. 105–121.

Carson, R., T. Groves, (2007), Incentive and informational properties of preference questions. *Environmental and Resource Economics* 37: 181–210.

Champ, P. and R. Bishop (2001), Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias. *Environmental and Resource Economics* 19: 383–402.

Champ, P., R. Bishop, T. Brown, and D. McCollum (1997), Using Donation Mechanisms to Value Nonuse Benefits From Public Goods. *Journal of Environmental Economics and Management* 33: 151–162.

Cook, J., D. Whittington, D. Canh, F. R. Johnson, and A. Nyamete (2007), Reliability of Stated Preferences for Cholera and Typhoid Vaccines with Time to Think in Hue Vietnam. *Economic Inquiry* 45: 100–114.

Creutzig, F., and D. He, (2009), Climate change mitigation and co-benefits of feasible transport demand policies in Beijing. *Transportation Research Part D: Transport and Environment* 14(2): 120–131.

Cummings, R., S. Elliot, G. Harrison, and J. Murphy (1997), Are Hypothetical Referenda Incentive Compatible. *Journal of Political Economy* 105: 609–621.

Cummings, R., G. Harrison and E. Rutstrom (1995), Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible. *American Economic Review* 85: 260–266.

Cummings, R. G., and L. O. Taylor (1999), Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. *American Economic Review* 89 (3): 649–665.

De-Magistris T., and S. Pascucci (2014), The effect of the solemn oath script in hypothetical choice experiment survey: A pilot study, *Economic Letters*, 123: 252–255.

Ehmke, M. D., J.L. Lusk, and J.A. List (2008), Is Hypothetical Bias a Universal Phenomenon? A Multinational Investigation. *Land Economics* 84 (3): 489–500.

Fifer, S., J. Rose, ., and S. Greaves, (2014), Hypothetical bias in stated choice experiments: Is it a problem? And if so, how do we deal with it? *Transportation Research Part A* 61: 164–177.

Frykblom, P. (1997), Hypothetical Question Modes and Real Willingness to Pay. *Journal of Environmental Economics and Management* 34: 275–287.

Hao, J., and L. Wang., (2005) Improving Urban Air Quality in China: Beijing Case Study. *Journal of the Air & Waste Management Association* 55(9): 1298–1305.

Hensher, D. A., J.M. Rose., and W.H. Greene, (2005), Applied choice analysis: A primer, *Cambridge University Press*.

Hensher, D. A. (2010), Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B* 44: 735–752.

Hultkrantz L., and S. Xue (2010), Mitigating hypothetical bias in value of time studies: Lab-experiment results, Working paper, *Orebro University*.

Isacsson G., (2007), The tradeoff between time and money: Is there a difference between real and hypothetical choices? Working Paper (2007;3) Swedish National Road and Transport Research Institute.

Jacquemet, N., R. V. Joule, S., Luchini, and J. F. Shogren (2013), Preference Elicitation under Oath, *Journal of Environmental Economics and Management* 65: 110–132.

Jacquemet, N., A. James, S. Luchini, and J. Shogren (2017), Referenda under oath, *Environmental and Resource Economics*, 67 (3): 479-504.

Jacquemet, N.,S. Luchini, J. Rosaz, and J. Shogren (2014), Truth-telling under Oath. Working paper, Documents de travail du Centre d'Economie de la Sorbonne 2015.68.

Joule R-V., F. Girandola, and F. Bernard (2007), How can people be induced to willingly change their behavior? The path from persuasive communication to binding communication. *Social Personality Psychology Compass* 1 (1): 493–505.

Li, Z., and D. Hensher (2012), Congestion charge and car use: A review of stated preferences and opinion studies and market monitoring evidence. *Transport Policy* 20: 47–61.

List, J. and C. Gallet (2001), What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values? *Environmental and Resource Economics* 20: 241–254.

List, J.A (2001), Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auction Experiments. *American Economic Review* 91: 1498–1507.

Murphy, J.G. Allen, T. Stevens, and D. Weatherhead (2005), A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics* 30: 313–325.

Qin B. and J. F. Shogren (2017), Is a promise as good as an oath? Evidence from hypothetical bias experiments in China, Working Paper, Xi'an Jiaotong University.

Scarpa, R., T. Gillbride, D. Campbell, and D. Hensher (2009), Modeling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics* 36: 151–17

Stevens T.H., M. Tabatabaei, and D. Lass (2013), Oaths and hypothetical bias. *Journal of Environmental Management*, 127: 135–141.

Swärdh, J.-E. (2008), "Hypothetical bias and certainty calibration in a value of time experiment" Working paper (2008: 8) in Swedish National Road & Transport Research Institute.

Vossler, C.A., M.F. Evans, (2009), Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management* 58(3): 338–345.

Westerdahl, D., X. Wang, X. Pan, and K.M. Zhang, (2009), Characterization of on-road vehicle emission factors and microenvironmental air quality in Beijing, China. *Atmospheric Environment* 43(3): 697–705.

Whittington, D., V.K. Smith, A. Okorafor, A. Okore, J.L. Liu, and A. McPhail (1992), Giving Respondents Time to Think in Contingent Valuation Studies: A Developing Country Application. *Journal of Environmental Economics and Management* 22: 205–25.

Yin, H. (2017). Value of time for private and public transport in Beijing, Unpublished working paper.

APPENDIX

Table A1. Probabilities of attending/ignoring attributes; results from an LCM model.

	Car		Public transport	
	No Oath	Oath	No oath	Oath
Consider all attributes	0.599*	0.591***	0	0
Ignoring morning time	0	0	0.446***	0.494***
Ignoring evening time	0	0	0	0.
Ignoring congestion charge	0.352***	0.335***		
Ignoring crowdedness			0.429***	0.356**
Ignoring fuel/ticket cost	0.049***	0.074***	0.125***	0.151**
No. obs.	315	357	377	365
McFadden Pseudo R^2	0.19	0.16	0.49	0.52